Metadata Evolution for NASA Earth Science Data Systems

MENDS Recommendation Report

A presentation to the

NASA Earth Science Data Systems Working Group

2011-11-06

Scope and Purpose of Presentation

- Origin and objectives of MENDS study
- Summary of MENDS recommendations report
- Activities since submission of report

Motivation

- NASA Earth Science Data Systems are heterogeneous, distributed and generate a diversity of data products
  - Metadata is the lifeblood of these systems
  - These systems employ many different data models
- Adopting ISO 19115 would:
  - Improve the usability and interoperability of NASA's data products
  - Help new NASA missions create data products with advanced quality, lineage and service access metadata
  - Help align NASA's systems with international partners
- NASA sought an enterprise-wide solution for ISO 19115 adoption that was:
  - Configurable
  - Evolvable
  - Allows for different levels and phases of adaptation
  - Involves both producer and user communities

Initiation of MENDS

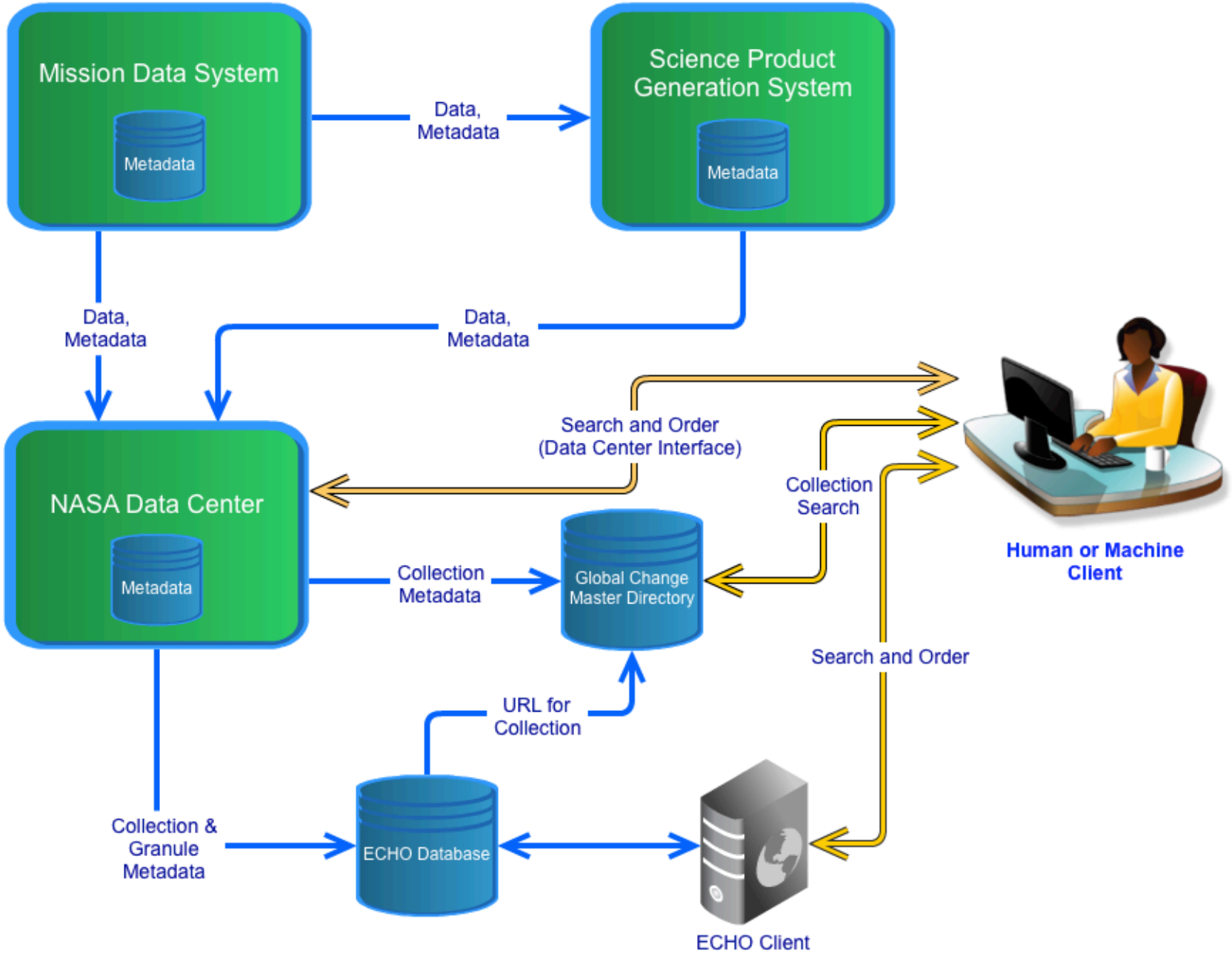
- In August 2010 NASA ESDIS Project assembled a team of stakeholders and experts from:
  - Decadal Survey missions,
  - Clearinghouses (ECHO and GCMD),
  - NASA Data centers and SIPS,

- And directed them to:
  - Study metadata needs and current practices,
  - Assess how ISO 19115 does or does not meet stakeholders' needs,
  - Determine optimal path for integration of existing data models with ISO 19115, and
  - Produce a recommendations report within 6 months.

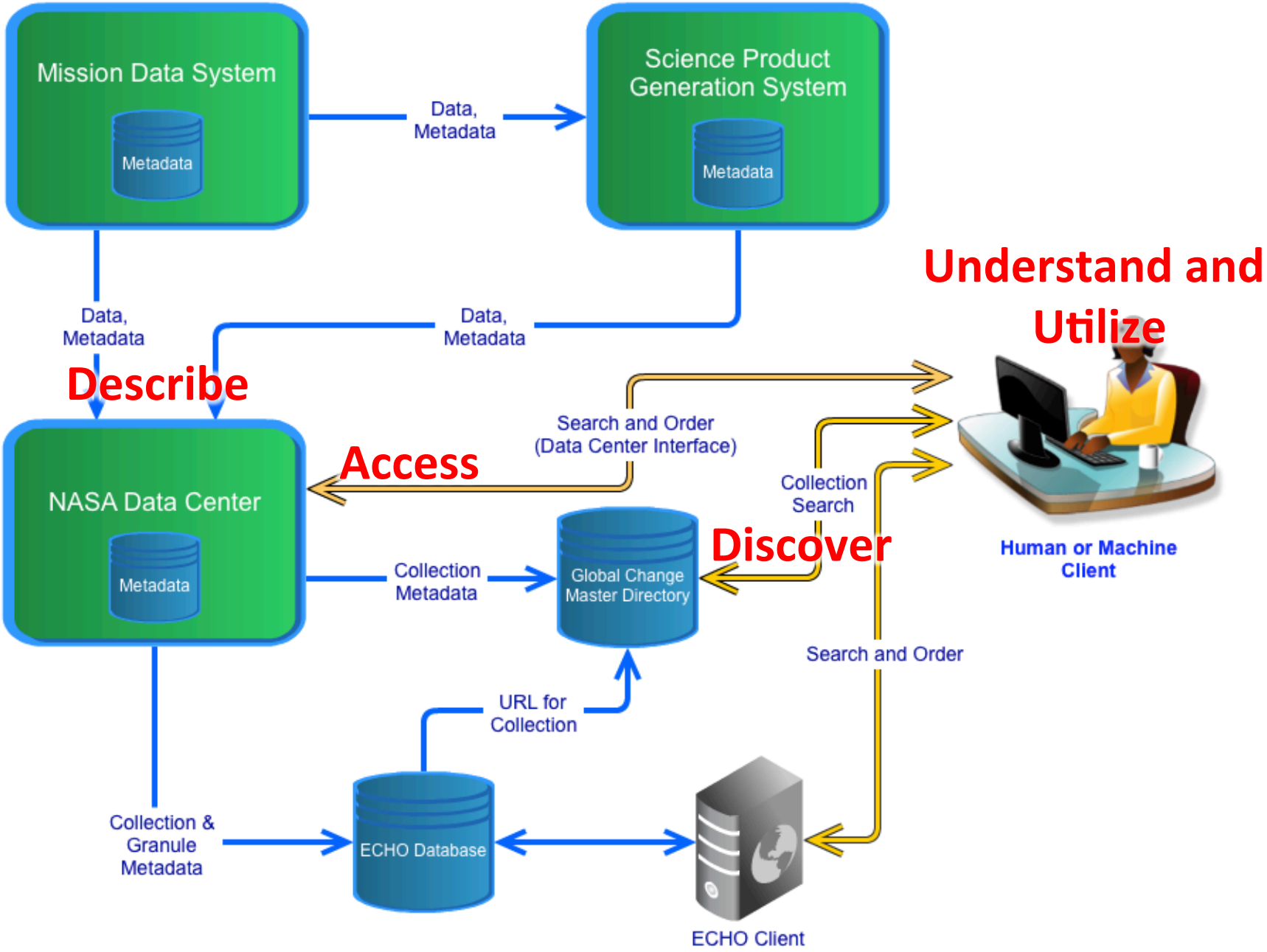
Methodology

- Study existing profiles of ISO 19115 and mappings from other standards to 19115
- Inventory current metadata needs and practices among NASA stakeholders
- Examine existing metadata flows between elements of NASA data systems
- Develop metadata usage scenarios

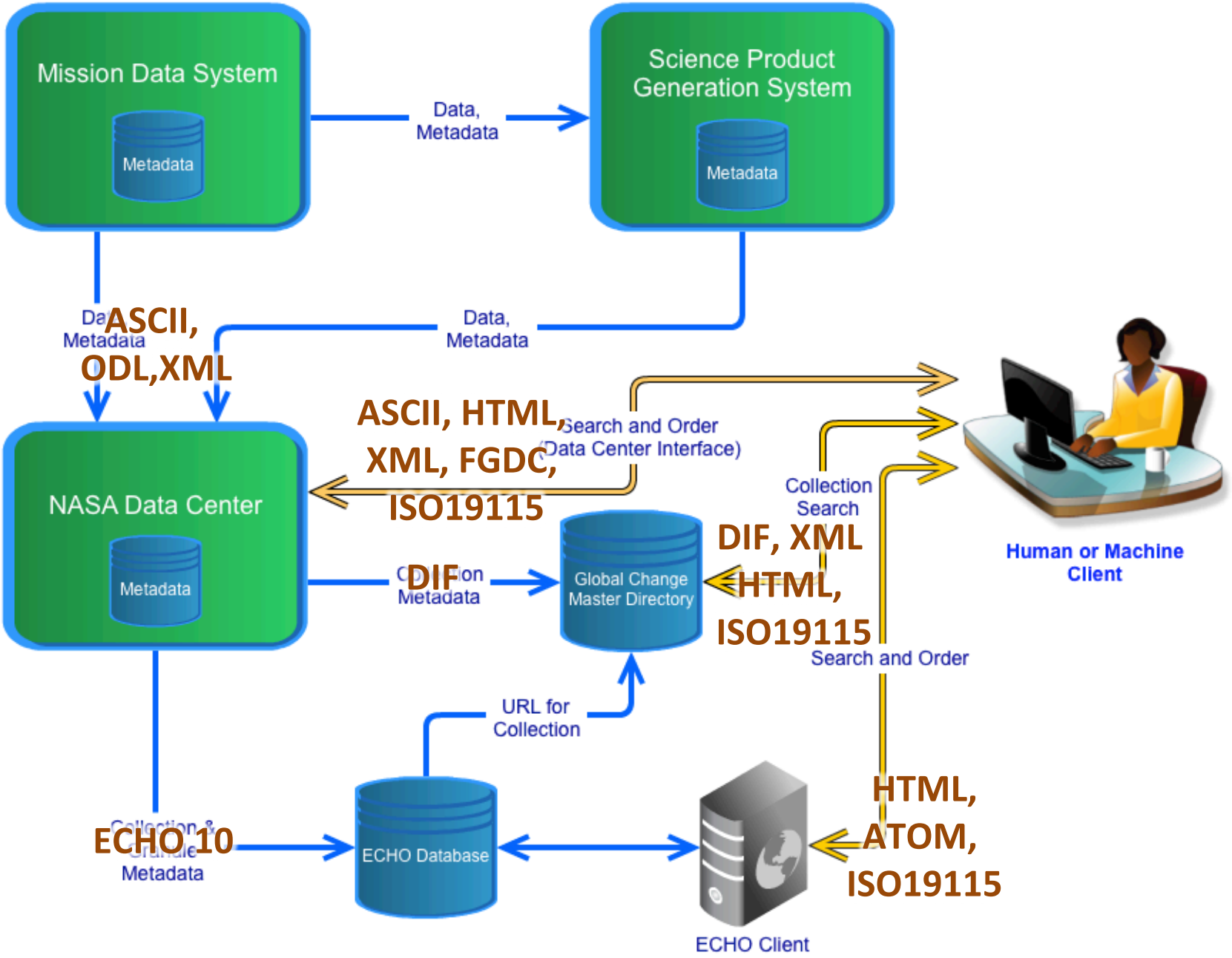
Metadata Flows in NASA Earth Science Data Systems



Some Functions of Metadata



Metadata Encodings and Formats Currently in Use



MENDS Process

- Phase I (August 2010 – January 2011)
  - Gain deeper knowledge of the 191** suite of standards
  - Examine existing metadata usage in EOSDIS
  - Collect stakeholder requirements and concerns
  - Analyze data and make recommendations

- Deliver technical paper to ESDIS
  - Description of benefits of ISO 19115+ over existing metadata models and processes
  - Assessment of impacts on existing systems and potential incremental costs for new systems
  - Proposed mechanisms for reducing cost and effort

Summary of Key Findings

- Adoption of ISO 19115+ is a logical step in the evolution of NASA Earth Science Data Systems
- Common usage across NASA systems of ISO 19115+ attributes and conventions for certain kinds of information is highly desirable
  - Argues for an extension, profile, or guidance on recommended practices
  - May need to be specific to observation type or science discipline
- Must avoid pitfalls of rushing to develop a profile
  - First gain experience implementing ISO 19115+ with NASA data products
  - Stay as close to original model as possible
  - Foster a community-driven process
  - Evolve the adoption of ISO 19115 in support of stakeholder needs
  - Learn from other implementations (NOAA, ESA, INSPIRE, etc.)
- Work to minimize impacts of ISO 19115 adoption on stakeholders
  - Support phased implementation by legacy systems
  - Provide tools for generation and validation

MENDS Team Recommendations

- NASA should establish conventions for the application of ISO 19115 to its Earth Science data products
  - Develop initial baseline schema as a starting point for early adopters
  - Important to gain experience implementing ISO 19115 before proposing a formal profile
  - Conventions may need to be specific to observation type or science discipline
  - Conventions should evolve through a community-driven process
  - Learn from/align with other implementations (NOAA, ESA, INSPIRE, ...)
- Work to minimize impacts of ISO 19115 adoption on stakeholders
  - NASA should avoid mandates whenever possible
  - Support phased implementation by legacy systems
  - Develop tools for generation and validation of metadata conforming to NASA conventions

MENDS Team Recommendations - 2

- Metadata is dynamic
  - It evolves as a product moves from production to archival and distribution, and as more is learned about it and services on it are enabled
- First step should be develop mappings and translators for existing repositories
  - Requiring little or no modifications to native data models
- NASA should track and document its implementations of ISO 19115, regardless of whether this leads to a formal profile or not
- The ISO 19115 adoption strategy should accommodate existing legacy systems while anticipating the needs of new participants and missions

Recommendations

- Eight overarching recommendations
- Targeted recommendations for
  - Evolution (8)
  - Decadal Survey Missions (4)
  - Data Centers (2)
  - ECHO (5)
  - ECS (3)
  - GCMD (2)
  - Implementation Strategies (4)

Overarching Recommendations

- ESDIS should adopt the ISO 19115 standard
- ESDIS should create NASA-specific conventions for use of ISO 19115
- The NASA ISO 19115 conventions should evolve in a formal manner, with the possibility of realizing a NASA ISO 19115 extension or profile
- ESDIS should put in place a mechanism for involvement of the broader stakeholder community in the evolution of NASA ISO 19115 conventions
- ESDIS should ensure that NASA stakeholders have adequate technical and financial support for their participation in the creation and adoption of the NASA ISO 19115 conventions
- ESDIS should generate a baseline schema for the NASA ISO 19115 conventions
- ESDIS should develop a comprehensive plan for the development and adoption of NASA ISO 19115 conventions
- NASA's heritage metadata repositories (ECS, ECHO and GCMD) should take specific actions to support the NASA ISO 19115 conventions

Recommendation IS-1

- “The MENDS team recommends that NASA stakeholders begin complying with the evolving NASA ISO 19115 conventions as their implementation plans dictate, and not wait until there is a complete set of NASA ISO 19115 conventions.”
- Contributing to the definition of the conventions comes first
- Determining compliance will require tools

Recommended Next Steps – Phase II

- Develop tools for early adopters that generate and read ISO metadata
- Stakeholder groups develop implementation plans
- Specify initial NASA ISO 19115 conventions to enable early adopters to begin their work
- Identify specific ISO elements for search and discovery
- Publish a preliminary ISO 19139-compliant schema definition for use with all NASA ESDIS datasets
- Institute ISO 19115 training

Recommended Next Steps – Phase III

- Establish procedures and a governance structure to guide the evolution of the NASA ISO 19115 conventions
- NASA ES-wide adoption of the initial version of a complete NASA ISO 19115 convention
- Refinement and wide release of ISO authoring and validation tools
- Refinement of ISO training
- Reassessment of stakeholders resource requirements for ISO adoption, based on better knowledge of the standard and initial experience of the early adopters

Community Engagement

- Involvement of the stakeholder community, both within and outside of ESDIS, is essential for success
- Coordination of NASA efforts with those of the larger Earth Science community will reduce the workload of NASA stakeholders
  - Establish and actively maintain a website that provides resources and hosts a forum for sharing implementation experiences
  - Track developments in other agencies and organizations and disseminate this information
- Formalize NASA's participation in ISO/TC211 to ensure that its standards evolve in a way that benefits NASA

Stakeholder Implementation Plans

- Each implementation plan will be based on the set of MENDS recommendations that ESDIS accepts. At the very least, the implementation plans must include:
  - A clear definition of the various training, design, implementation and test steps required to incorporate ISO metadata into the stakeholder's system,
  - An estimate of the amount of time and effort required to accomplish each task, and
  - An estimate of any other resources that might be required to accomplish each task
- Being done by ECHO, GCMD, SMAP and some DAACs

Some Open Issues

- Placement of metadata inside data files versus having separate metadata files, tracked with persistent URIs, inside the data files, and what the URIs would be;
- Which ISO 19115 metadata elements should be used for existing NASA metadata attributes when multiple ISO 19115 options are available;
- For certain NASA metadata attributes does it make more sense to use another standard, e.g. SensorML, or another standard in the ISO 19115 family, such as ISO 19130?
- These and other issues to be addressed during MENDS Phase II, supported by further research and testing.

Review and Endorsement

- The MENDS report has yet to be officially endorsed, but is being released with the proviso
  - "**DRAFT Version Under Project Review. For Information Only. Any questions, please contact Andrew Mitchell, Andrew.E.Mitchell@nasa.gov**"

Summary

- ISO19115 provides tremendous depth and flexibility in documenting all facets of a data product
  - This complexity necessitates substantial investments to fully understand and work with the standard
  - Capable tools can greatly reduce the effort involved in designing, generating and validating ISO metadata
  - No one should have to read ISO specs if they don't want to!
- NASA ESDIS has the opportunity to play a leading role in the application of ISO 19115+ standards to remote sensing data
- Demonstrating benefits is better than mandating conformance

The MENDS Team

- Alicia Aleman, Tyler Stevens *NASA GCMD*
- Ed Armstrong *NASA/JPL PO.DAAC*
- Tammy Beaty *NASA ORNL DAAC*
- Steve Browdy *OMS Tech*
- Matt Cechini, Dan Pilone *NASA ECHO*
- Helen Conover *NASA GHRC/ UAH*
- Patrick Denny, Rudi Gens *Alaska Satellite Facility*
- Joe Glassy *U. Montana*
- Ted Habermann *NOAA/NGDC*
- Siri Jodha S Khalsa *NSIDC DAAC*
- Christopher Lynnes *NASA GES DISC*
- Andrew Mitchell *NASA/ESDIS*
- Jon Pals, Bob Hartranft *Raytheon EED*
- Barry Weiss *NASA/JPL, SMAP*
- Wenli Yang *NASA/GSFC, GMU*